DOCUMENT RESUME

ED 279 684 TM 870 071

AUTHOR Jones, Lyle V.
TITLE The Future Assessment of Educational Progress: Specifying Background Variables and Subpopulations.
PUB DATE 86
NOTE 19p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Biographical Inventories; *Community Characteristics; Comparative Analysis; *Educational Assessment; Elementary Secondary Education; *National Surveys; *Sampling; State Surveys; *Student Characteristics; Testing Programs; *Test Results
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

It has been proposed that the National Assessment of Educational Progress (NAEP) increase its efforts to gather and analyze data on student background, teachers, households, and schools. Adoption of this proposal would result in the inclusion in NAEP of additional background variables and subpopulations, to be defined by states. NAEP specification of subpopulations has allowed efficient sampling, examination of various subpopulations' achievement levels, and interesting statistical comparisons. Traditional NAEP subgroups have involved community characteristics (region, size, and type) and student characteristics (age; sex; ethnic group; parental education; and time spent on activities such as television, reading, and homework). It is desirable to maintain those reporting categories judged to be policy relevant, whether or not they have been consistent with national achievement trends. It is difficult to define and measure the most important policy relevant variables; prior research should support their validity as reliable indicators of the intended characteristics or process. Future reporting categories should allow meaningful comparisons for 1969-1987. Central control of data is suggested. Suggestions for state sampling and reporting of results are also presented. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

4

ED279684

The Future Assessment of Educational Progress:

Specifying Background Variables and Subpopulations

Lyle V. Jones

University of North Carolina at Chapel Hill

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L.V. Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

TM 870 071

THE FUTURE ASSESSMENT OF EDUCATIONAL PROGRESS: SPECIFYING BACKGROUND VARIABLES AND SUBPOPULATIONS

Lyle V. Jones

The University of North Carolina at Chapel Hill

A primary goal of the National Assessment of Educational Progress (NAEP) has been that of tracking changes in school achievement over time for the nation's 9-, 13-, and 17-year-old children. Since its inception, another goal of NAEP has been to report achievement results at each age by population subgroups, comparing subgroups with each other and with the nation as a whole for both average performance levels and changes in those levels.

Secretary of Education William J. Bennett, called recently for the planning and development of improved methods for the national assessment of American education. Bennett set forth several principles, including that "states should be encouraged to participate more fully in the National Assessment program to enable them readily to guage the performance of their students against the performance of the nation as a whole and of other states," and that "NAEP should continue to increase the emphasis on gathering pertinent background information about students, teachers, households and schools and should analyze its test data in relation to such information" (Bennett, 1985). These principles invite the inclusion in NAEP of additional subpopulations, to be defined by state, and of additional background variables. If adopted, these principles will have profound effects on national assessment.

Before considering the implications of a NAEP design that allows state-by-state comparisons and increases the emphasis on background information, it is of interest to review the background information that has been collected and employed by NAEP and to consider the utility of that information, not only for defining subgroups as units for reporting achievement results, but for other purposes as well.

In NAEP, the specification of subpopulations has served one or more of three distinct purposes:

(1) A more efficient sampling design may be derived by establishing differential sampling rates for different subpopulations (see National Assessment of Educational Progress, 1974). In the context of (2), below, when it is desired to report findings for a subpopulation that constitutes only a small portion of the total population, it becomes especially important to assure adequate sampling of that subpopulation.

(2) The average performance levels of some subpopulations are of intrinsic interest, and achievement results can be reported separately for those population subgroups.

(3) It sometimes is of interest to ask "what if" questions of assessment data, as in National Assessment of Educational Progress (1973). Thus, one might ask of 1980 data, what if the distribution of students by size and type of community had been the same as it had been in 1970? How then would 1980 national achievement averages for 13-year-olds be estimated to have compared with those for 1970? More importantly, how much of the apparent change in achievement between 1970 and 1980 might be

associated with a shift in the distribution of students over categories of size and type of community? An answer to such a question is suggested by using statistical methods for adjusting 1980 data to improve their comparability with respect to size and type of community to 1970 data, but only if data for both years can be disaggragated by community size and type.

"What if" questions may become of greater interest if assessment results are available by state as well as for the nation as a whole. In addition to the interest in average achievement in a state relative to the national average, it also is of interest to ask what if the population characteristics of that state more closely resembled those of the nation (or of another state), and how then would achievement in the state compare with that of the nation (or of the other state).

The Traditional NAEP Subgroups

Since its inception, NAEP has regularly collected background information that has allowed NAEP reports to include achievement results presented separately for a number of subpopulations. The primary background variables have been of two kinds, representing community characteristics and student characteristics.

Grouping by Community Characteristics

Region. The 51 "states" (including the District of Columbia) are organized into four regions that have been used in the sampling design and that also have served as reporting categories for NAEP results, Northeast (12 states), Southeast (12 states), Central (12 states), and West (15 states).

Size of Community (SOC). NAEP has employed, both in the sampling design and as reporting categories, four SOC categories: Big Cities, all students attending school within the city limits of a city with a 1970 population of at least 200,000; Fringes around Big Cities, all students in the same urbanized area as a Big City, but outside the city limits; Medium Cities, all students attending school in cities with populations between 25,000 and 200,000 who are not in the Fringes category; Small Places, students attending schools in communities with a population less than 25,000 who are not in the Fringes category.

Type of Community (TOC). Subsets of students from selected SOC categories have been singled out to be reported in TOC categories (referred to in some NAEP reports as STOC categories, for Size and Type of Community). The TOC categories are: Advantaged Urban, students from Big Cities or Fringes with a high proportion of citizens in professional or managerial employment; Disadvantaged Urban, students from Big Cities or Fringes with a relatively high proportion of citizens on welfare or not regularly employed; Rural, students attending schools in communities with a population less than 10,000 where many residents are farmers or farm workers; Other, all remaining students.

Grouping by Student Characteristics

Age. A major reporting and design factor for NAEP since its inception has been the age of the student. NAEP has assessed students at ages 9, 13, and 17. In the "new" NAEP design, students are sampled by grade (grades 4, 8, and 12) as

well as by age; some users of NAEP findings prefer viewing results by grade rather than by age, since grade constitutes a traditional reporting unit in the schools. As noted by Messick, Beaton & Lord (1983), a disadvantage of reporting by grade is that the distribution over grade levels of students at a given age may vary from time to time (and from state to state) as a direct consequence of policies that determine the age at which schooling begins and the retention of students in grade or their promotion from grade to grade. Average performance levels by grade, consequently, might be controlled in part by the adoption of different schooling policies, thereby influencing assessment results and affecting their comparability from one time to another (or from one state to others). With these considerations in mind, Messick et al. (1983) wisely advise that results be collected and reported both by age and by grade.

Sex. All NAEP assessments have provided for separate reporting for female and male students.

Ethnicity. The NAEP specification of race/ethnicity has varied from time to time. Since NAEP Year 1 (1969-70), results have been able to be separately reported for black students and white students. For some assessments, results also have been reported for Hispanic students (included as white students when only two categories were employed). The small numbers of native American students and students of Asian origin typically have precluded the separate reporting of NAEP results for students in these categories.

Parental Education. NAEP results can be categorized by four levels of parental education, for students who report the highest level of education of either parent to have been (1) the 8th grade, (2) higher than the 8th grade but lower than graduation from high school, (3) high school graduation, and (4) some formal education beyond high school. (In National Assessment of Educational Progress, 1985, categories (1) and (2) were collapsed when results were reported.)

Other Classifications. Over the history of the NAEP project, results sometimes have been reported by other sets of categories, often derived from a Student Questionnaire. Thus, in National Assessment of Educational Progress (1985), results were presented by student reports of Number of Reading Materials in the Home (few, some, or many), Hours per Week of Television Viewing (0-2, 3-5, 6 or more), and Hours per Night of Homework (none assigned, did not do it, less than 1, 1-2, more than 2). In National Assessment of Educational Progress (1981), results were reported by four quartiles of Achievement Class, determined by student performance relative to all other students who were administered the same set of reading exercises. Clearly, any variable that is available for student respondents may be employed to identify sugroups for reporting.

Issues Pertinent to the Selection of Subgroups

Some subpopulations are likely to be of greater interest than others to policy makers, e.g., groups defined by state of residence, socioeconomic class, type of community, participation in special programs or curricula, or time spent studying.

It might be argued that grouping students into categories, each of which in the past has yielded achievement trends like those of the nation as a whole, is essentially a waste of effort. By this argument, the most interesting groupings would be those for which NAEP findings have provided "surprises," in the form of trends that differ from national trends. Candidates include Ethnicity (achievement gains have been greater for black and Hispanic students than for white students), Region (gains for the Southeast have exceeded those for other regions, especially at age 9), and Parental Education (higher gains have been found for children whose parents failed to complete high school than for children whose parents completed high school and/or attended college). Of course, categories that have failed to provide differential trends in achievement change in the past cannot be guaranteed to yield results in the future that track the national results. Therefore, it would seem desirable to maintain as potential reporting categories those that are judged to be policy relevant, whether or not they have proven in past NAEP reports to be parallel to national achievement trends.

One set of categories that may be expected to retain policy relevance, and that also has yielded interesting achievement trends, is that defined by student ethnicity. Trends in achievement for black and Hispanic students have differed from trends for the nation, e.g., Burton & Jones (1982), Jones (1984), National Assessment of Educational Progress (1985), and, because of special concern for the quality of education that is provided to minority students, considerable attention has been paid to

achievement trends as reported by ethnic group. It should be recognized, however, that reporting by categories such as those of ethnicity invites misinterpretation to the extent that the reporting categories are assumed to be causally related to achievement. By way of illustration, average achievement levels for black and Hispanic children are found to be lower than for white children. But these subpopulations differ from each other in many ways other than by ethnic category, and some of these ways are highly associated with achievement levels within ethnic group, e.g., the proportions of students from poverty households, the proportions from non-English-speaking homes, the frequency and quality of preschool educational experiences, the distribution of parental education, etc. Many factors other than ethnicity per se are likely causal agents.

A recurrent problem in selecting reporting categories (subpopulations) in survey research is that those with the greatest potential policy interest often are those that are hardest to measure. There is a temptation, then, to accept -- as a substitute for the desired variable -- an unreliable indicant of that variable, producing categories that are fuzzy, at best, and that may be found to be associated with little systematic variability in performance, even when the underlying but hard-to-measure variable is intimately related to performance. At least as troubling, if the categories are used for the purpose of statistical adjustment, is the failure to explicitly recognize that they will necessarily be ineffective whereas adjustment by the underlying variable of interest would have been highly effective.

Variables to be employed in a national assessment should be selected from variables that have been validated in prior research as reliable indicators of the characteristics or process that is intended to be represented. Lacking sufficient evidence from past research, special studies should be conducted to establish the validity of the variables before they are employed in a national assessment.

The Importance of Continuity with NAEP

In planning for the future assessment of educational progress, it is important that the project be designed to provide meaningful comparisons with findings from NAEP for the period 1969-1987. This will entail a variety of constraints, including some constraints on the specifications of subgroups, in order that achievement changes may be displayed for major subgroups that have been consistently designated in NAEP reports, most notably for Region, Size and Type of Community, Sex, Ethnicity, and Parental Education. These reporting categories have been maintained throughout the history of NAEP, and because each of them remains of intrinsic interest, it is recommended that they also be routinely maintained as a minimal set for future national educational assessments. (Were results to be reported separately for all states, it might no longer be important to report by Region. However, especially if results are available for some states and not for others, continued reporting by Region will serve as important points of reference for states within region.)

Perhaps the most critical aspect of NAEP, for assuring comparability with past results and comparability of results among the states, is the maintenance of central control of sampling, data collection, analysis and reporting. Quality control is essential to assure comparability of findings and will continue to require that full responsibility for these facets reside with a national assessment staff. This requirement will loom larger in importance when results are reported by state, as pressure will mount to enhance the efficiency of data collection by delegating administrative authority regionally or by state. But as state-by-state reporting of results becomes possible, personnel in a state will develop keen interest in the achievement outcome. A fundamental tenet of research design requires that responsibility for data collection be vested in neutral parties rather than in interested parties, thereby adding strength to the demand for central control of data collection.

Costs of Reporting by Subgroup in State-by-State Assessment

For national subgroup results to be useful, it is necessary that data be collected in a way that assures that the sizes of standard errors of mean achievement by subgroup remain modest. This requires adequate sample sizes for the subgroups, and may require the oversampling of groups that make up only small parts of the total population. In an assessment designed to allow state-by-state comparisons, the cost of adequate sampling of subgroups within every state is likely to be formidible, especially because some subpopulations will be sparsely represented in a given state. Whatever the choices of

background characteristics to be collected for the purpose of defining reporting groups in the national assessment, a choice might be provided to participant states, either to select or not to select a sampling design adequate for reporting state results by some or all of the subgroups, with the choice expected to influence the cost of participating in the study. Another alternative to mainaining subgroup specifications by state in a national assessment program is to perform special state assessments that have been designed to detect achievement differences by subgroup, rather than to meet that need through a national assessment program. The alternatives should be carefully evaluated with respect to both cost and expected utility of findings before deciding what subpopulations to include by state in a national assessment.

Pros and Cons of Adjusting State Assessment Results for Disproportionate Representation over Subgroups

Let us assume that a plan for national assessment is adopted in which results are available by state, at least for some states, and also that results within some states are available by selected population subgroupings, e.g, by Sex, Ethnicity, Parental Education, and TOC, and that these dimensions have been adequately included in the state sampling frame. What are the arguments, then, for and against the reporting of results after adjustment for disproportionate representation over subgroups?

Reporting without Adjustment. Average achievement levels will be available by whole state, and it is assumed that these are to be reported, regardless of the decision as to whether or not to report results after adjusting by student and community background characteristics. When results are reported, it would be desirable also to report, both for the state and for the nation, the distribution of students over the relevant categories of the background variables. This would provide a basis for an informal and subjective "correction" of state results for possible disproportionate representation of students in a state over the categories of those variables.

Reporting by Subcategories. Rather than employing statistical methods of data adjustment, results by state might be presented in a stratified format, separately by category of student background variables. This would allow a focus not only on overall state-by-state results, but on state-by-state differences for comparable groups of students. Reporting results by category of background variables can either be a substitute for reporting after adjustment, or can adopted as a complementary approach.

Reporting Results after Adjusting for Disproportionate Representation over Subgroups. NAEP results from Year 1 were adjusted and presented in the form of estimates of achievement relatively uncontaminated by interfering variables, using the technique called balancing (National Assessment of Educational Progress, 1973). Wiley (1976b) discusses the use of both direct and indirect standardization as techniques to achieve similar objectives,.and Wiley (1976a) illustrates the use of analysis of

covariance to isolate effects on achievemant of "quantity of schooling;" the alternative techniques of balancing, direct standardization, indirect standardization, and analysis of covariance have been examined by Maxwell & Jones (1976), where the choice among them was shown to be a complex function of sample size, concern about bias, concern about standard errors, and goodness of fit of assumptions to population features.

Methods of statistical adjustment are powerful tools for suggesting background variables that may be influencing levels of school achievement. The results from these methods, however, typically allow a multiplicity of interpretations, choices among which depend not only on statistics but also on theory, the understanding of educational processes, and an intimate acquaintance with the measures that are employed (as well as with those that might have been but were not included in the study). For these and related reasons, the production of reports based on statistical adjustments of data are unlikely to be successful if produced under severe time constraints. The complexity of the task typically demands time for alternative analyses to be conducted and for alternative interpretations to be developed and carefully evaluated prior to the release of a report of findings and conclusions. This demand may appear to conflict sharply with one of Secretary Bennett's principles, that "NAEP should markedly improve its capacity to get all data speedily analyzed and promptly disseminated" (Bennett, 1985). The principle is important. The conflict should be avoided by asking that the basic, descriptive NAEP findings be promptly

reported, while accepting delay as part of the cost of more detailed and more complex analyses.

Further Suggestions for the Selection of Background Variables

For selecting variables that correlate with educational achievemant, guidance abounds in the literature of educational research. Some of that guidance has been distilled by a National Research Council Committee on Indicators of Precollege Science and Mathematics Education and is set forth in their preliminary report (Raizen & Jones, 1985). (A final Committee report is nearing completion.) The following discussion owes much to the work of this Committee.

Community Variables. It is recommended that Region, SOC, and TOC be retained, and that State be added.

Student Variables. Age, Sex, Ethnicity, and Parental Education should be retained. Amount of Homework, Reading Materials in the Home, and Amount of Television Watching, as reported by National Assessment of Educational Progress (1985), should be considered for regular inclusion. When assessing a particular subject matter, efforts should be made to include information on the number of semesters of relevant courses taken by the student, information about student interests in, attitudes about, and motivation to learn that subject, and information about out-of-school activities related to learning, e.g., for science, attendance at science museums, watching science programs on TV, and the like. Extent of library use, and extent of use of laboratory equipment and computers, also should be considered.

School, Teacher, and Curriculum Variables. It must be recognized that the achievement of a student at any particular time is a cumulative function of years of learning, both in school and out, and that school, teacher, or curriculum characteristics at that time should not be expected to play a large role in influencing student achievement. Even so, some school, teacher, and curriculum variables should be considered. Mounting evidence suggests the importance of teacher expectation on student achievement. (Expectations of principals, parents, and the students themselves also are influential.) Evidence also supports the importance of the constructive use of classroom time. Measures of teacher expectation and use of classroom time are candidates for inclusion.

Teacher characteristics such as sex, race, highest degree, undergraduate major, and salary, have not been shown consistently to be related to effective teaching, and for that reason are not likely to be good candidates for inclusion.

Curriculum coverage and curriculum quality are relevant features of schooling, but are so difficult to measure that it would seem wiser to study their influence by performing research with that limited goal rather than to try to include them in a national assessment.

Many research projects in progress, including longitudinal studies sponsored by the Department of Education, have among their objectives the detection of process variables that are related to student achievement. Results from those projects should be systematically monitored for additional candidate variables for national assessment.

REFERENCES

Bennett, W. J. Improving our ability to assess educational progress. Washington, D.C.: U. S. Department of Education, unpublished statement.

Burton, N. W. & Jones, L. V. (1982). Recent trends in achievement levels of black and white youth. Educational Researcher, 11, 10-14.

Jones, L. V. (1984). White-black achievement differences: The narrowing gap. American Psychologist, 39, 1207-1213.

Maxwell, S. E. & Jones, L. V. (1976). Female and male admission to graduate school: An illustrative inquiry. Journal of Educational Statistics, 1, 1-37.

Messick. S., Beaton, A. & Lord, F. (1973). National Assessment of Educational Progress Reconsidered: A new design for a new era. Princeton: National Assessment of Educational Progress.

National Assessment of Educational Progress (1973). Report no. 7, science group results, B. Washington D.C.: U. S. Government Printing Office.

National Assessment of Educational Progress (1974). The National Assessment approach to sampling. Denver: National Assessment of Educational Progress.

National Assessment of Educational Progress (1981). Three National Assessments of reading: Changes in performance, 1970-1980. Denver: Education Commission of the States.

National Assessment of Educational Progress (1985). The reading report card. Princeton: Educational Testing Service.

Raizen, S. A. and Jones, L. V. (Eds). (1985). Indicators of precollege education in science and mathematics: A preliminary review. Washington, D. C.: National Academy Press.

Wiley, D. E. (1976a). Another hour, another day: Quantity of schooling, a potent path for policy. In W. H. Sewell, R. M. Hauser, & D. L. Featherman (Eds.), Schooling and achievement in American society. New York: Academic Press.

Wiley, D. E. (1976b). Approximations to ceteris paribus: Data adjustment in educational research. In W. H. Sewell, R. M. Hauser, & D. L. Featherman (Eds.), Schooling and achievement in American society. New York: Academic Press.